信号检测论与贝叶斯决策理论的关系

胡啸

（北京师范大学心理学部，北京 100875）

摘　要：信号检测论被广泛用于解释个体在不同类型认知任务中的决策过程。然而，经典信号检测论的重要不足之处在于难以进一步解释个体设置报告标准的过程对应怎样的内在心理机制。本文从贝叶斯决策理论的视角出发，深入探讨了个体在信号检测任务中的决策规则。本文首先基于贝叶斯定理，介绍了贝叶斯决策理论的基本观点。随后，本文探讨了贝叶斯决策理论如何解释理想观察者的决策规则，以及在实际的信号检测任务中个体的决策结果与理想观察者之间的偏离。接下来，本文探讨了经典信号检测论和贝叶斯决策理论在不等方差信号检测模型中的差异。最后，本文简要介绍了支持贝叶斯决策理论的实证研究证据。

关键词：信号检测论；贝叶斯决策理论；先验概率；似然函数；报告标准

1　引言

信号检测论(signal detection theory, SDT)是实验心理学研究中应用最为广泛的计算模型之一。自从心理学家 John A. Swets 及其合作者首次将信号检测论系统性地引入心理学领域以来(Green & Swets, 1966; Tanner & Swets, 1954)，研究者们已经广泛地应用信号检测论模型来解释知觉、记忆、推理等不同心理过程的内在机制(Banks, 1970; Mamassian, 2016; Wixted, 2020)。在心理学论文数据库 PsycArticles 和 PsycInfo 中，以“signal detection theory”为关键词可以检索出超过 4000 篇文献，且在 2020 至 2022 这三年间发表的文献就达到了 500 篇，

基金项目：国家自然科学基金项目（32200841; 32171045）。
通讯作者：胡啸，E-mail: xiaohu@bnu.edu.cn。

足以证明信号检测论不仅在心理学研究的历史上占有重要地位，同时在今天仍然具有旺盛的生命力。

在基于信号检测论的心理学实验设计中，研究者会向被试呈现两类不同的刺激，其中一类刺激被称为“信号”(signal)，而另一类刺激被称为“噪音”(noise)。被试的任务是判断哪些刺激属于信号，哪些刺激属于噪音(Wickens, 2001)。比如，在一个听觉任务中，被试听到的刺激可能只包含白噪音，也可能在白噪音的基础上叠加了一个特定声调的声音作为信号；被试需要判断信号在哪些试次(trial)当中出现(Egan et al., 1959)。又如，在再认记忆任务中，被试需要首先学习并记忆一系列词语，随后研究者会向被试呈现一些学过的“旧词”作为信号，和一些没学过的“新词”作为噪音；被试需要判断每个词语是新词还是旧词(Mickes et al., 2007; Wixted, 2007)。通常情况下，信号检测论会假设信号和噪音的刺激强度分别符合一个正态分布，且信号分布的均值高于噪音分布；信号分布和噪音分布均值之差被称为辨别力指数(d’)，它反映了个体辨别信号和噪音的能力，d’越高则个体的辨别力越强(Wickens, 2001)。

信号检测论关注的一个重要问题是，个体如何决策哪些刺激属于信号，哪些刺激属于噪音。经典的信号检测论观点认为，个体会直接在刺激强度的数轴上设置报告标准 C，如果当前刺激的强度高于 C，则个体判断当前刺激为信号，反之判断为噪音(Wixted, 2020)。图 1 给出了信号检测论模型的一个示例。该模型中噪音分布的均值为 0，信号分布的均值为 d’， 两

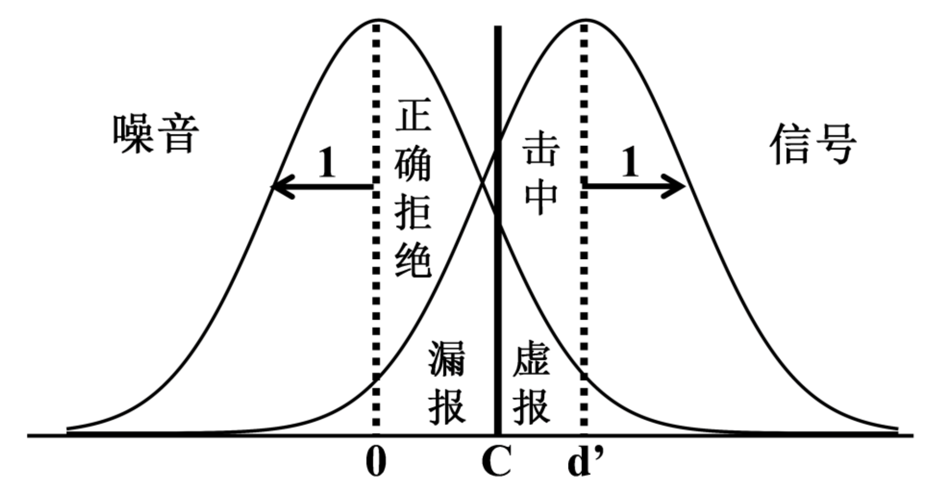


图 1 信号检测论模型示例（图中的模型为等方差信号检测模型）

个分布的标准差均为 1；当某个刺激的强度高于 C 时，个体会将其判断为信号，反之则判断为噪音。在信号检测任务中，个体的作答结果可以被分为四种类型：击中（将信号刺激判断为信号）、虚报（将噪音刺激判断为信号）、漏报（将信号刺激判断为噪音）和正确拒绝（将噪音刺激判断为噪音）(Wickens, 2001)。四类作答结果出现的概率由代表信号或噪音的正态分布在 C 的左侧或右侧的面积大小来表示（见图 1）。

信号检测论的优点在于，它可以从个体的作答正确率数据中，分离出辨别力指数 d’和报告标准 C 这两个指标，使得研究者可以分别探讨任务难度（即 d’）和个体的反应偏向（即 C）如何影响个体最终的作答结果(Wickens, 2001)。然而，经典信号检测论框架（即直接使用 C 来反映个体的决策判断标准的信号检测论模型）也存在一个重要的不足之处：它难以深入解释个体在信号检测任务中完成决策的过程对应怎样的内在心理机制。虽然经典信号检测论简单地假设个体会直接将当前刺激的强度与一个特定的判断标准 C 在数轴上的位置进行比较，并基于比较的结果来完成决策，但它通常难以进一步解释个体为何会把报告标准 C 设置在数轴上的特定位置，以及对于不同个体而言或在不同实验条件下 C 的位置产生差异的内在原因(Glanzer et al., 2009, 2019)。

事实上，早在信号检测论提出之初，Swets 及其合作者就开始使用贝叶斯决策理论(Bayesian decision theory, BDT)的观点来解释个体在信号检测任务中的决策过程(Green & Swets, 1966)。近年来，贝叶斯决策理论与信号检测论之间的关系开始受到越来越多的研究者的重视(Fleming & Daw, 2017; Glanzer et al., 2019; Maloney & Zhang, 2010)。贝叶斯决策理论认为个体会基于观测到的当前刺激强度，通过一个贝叶斯推理过程来决策当前刺激是信号还是噪音(Fleming & Daw, 2017; Maloney & Zhang, 2010; Pouget et al., 2016)。和经典信号检测论相比，贝叶斯决策理论为个体如何在信号检测任务中完成决策过程给出了更加深入的理论解释(Glanzer et al., 2019; Lau, 2007)。

本文将从贝叶斯决策理论的视角，深入探讨信号检测论框架下个体的决策规则。本文将首先介绍贝叶斯决策理论的基本观点，即个体在信号检测任务中如何通过一个贝叶斯推理过程完成决策。随后，本文将阐述信号检测论当中的“理想观察者”(ideal observer)问题，并在等方差(equal variance)信号检测模型的框架下，介绍贝叶斯决策理论如何解释实际研究中的被试与理想观察者之间的偏离。接下来，本文将讨论不等方差(unequal variance)的信号检测模型，并阐述贝叶斯决策理论与经典信号检测论在该模型中存在的差异。最后，本文将介绍支持贝叶斯决策理论的实证研究证据。

2　贝叶斯决策理论

贝叶斯决策理论的基础是贝叶斯定理(Bayes' theorem)，它来源于对联合概率(joint probability)的计算公式(胡传鹏　等, 2018)。假设有两个事件——A 和 B，则 A 和 B 同时发生的概率被称为联合概率，记为 P (A, B)。联合概率的计算公式可以被写为如下形式：

$$P(A,B) = P(B|A)P(A) \tag{1}$$

公式(1)的含义为：A 和 B 同时发生的概率，等于 A 发生的概率 P (A)乘以在 A 发生的条件下 B 发生的概率 P (B | A)。其中，P (A)又被称为边缘概率(marginal probability)，它指的是在不考虑其他事件的发生结果的情况下，某一事件发生的概率；而 P (B | A)又被称为条件概率(conditional probability)，它指的是在确定某一个事件（如 A）已经发生的条件下，另一个事件（如 B）发生的概率。实际上，A 和 B 同时发生的联合概率，还可以被写为另一种形式，即等于 B 发生的边缘概率 P (B)乘以在 B 发生的条件下 A 发生的条件概率 P (A | B)：

$$P(A,B) = P(A|B)P(B) \tag{2}$$

基于以上的公式(1)和(2)，可以得出如下的公式(3)：

$$P(A|B)P(B) = P(B|A)P(A) \tag{3}$$

公式(3)通常会被写为如下的形式：

$$P(A|B)=\frac{P(B|A)P(A)}{P(B)} \quad (4)$$

公式(4)即为贝叶斯定理的一般形式。贝叶斯定理的提出起初只是为了刻画两个事件的边缘概率和条件概率之间满足的数学关系。然而，心理学家很快意识到，贝叶斯定理可以被用来刻画人类的推理与决策过程(Kersten et al., 2004; Lau, 2007; Wickens, 2001)。比如，公式(4)可以被用来解释个体如何通过对事件 B 进行观测，来推理事件 A 的发生概率。此时，公式(4)中的 P (A)被称为先验概率(prior probability)，它指的是在观测到事件 B 之前，个体对于事件 A 发生概率的先验认识。P (A | B)被称为后验概率(posterior probability)，它反映了个体在观测到事件 B 的发生结果之后对事件 A 发生概率的认识。个体在贝叶斯推理过程中，会根据对事件 B 的观测结果来更新自身对事件 A 发生概率的认识，即从 P (A)更新为 P (A | B)。另一方面，P (B | A)被称为似然函数(likelihood function)，它反映了在给定事件 A 的情况下，事件 B 发生的可能性。在个体对事件 A 发生概率进行更新的过程中，先验概率和似然函数都起到了重要的作用：一方面，当 A 发生的先验概率值越高时，其后验概率值也会越高；另一方面，在假定事件 A 发生的情况下，如果事件 B 发生的概率越高（即似然函数值越大），则个体在观测到事件 B 之后也会认为事件 A 成立的后验概率越高。

根据贝叶斯决策理论的观点，在信号检测任务中，个体会观测到每个刺激的强度（记为 *x*），并基于 *x* 的值来完成一个贝叶斯推理过程，以推理当前刺激属于信号或噪音的后验概率(Burgess, 1985; Fleming & Daw, 2017; Wickens, 2001)。下面的公式给出了个体如何基于刺激强度 *x* 对当前刺激属于信号（记为 S）的概率进行更新：

$$P(S|x)=\frac{P(x|S)P(S)}{P(x)} \quad (5)$$

公式(5)意味着，在观测到当前刺激之前，个体对当前刺激属于信号的概率存在一个先验的认识，反映为先验概率 P (S)。而在观测到当前刺激的强度 *x* 之后，被试将当前刺激属于

信号的概率更新为后验概率 P (S | x)。同理，个体基于刺激强度 x 对当前刺激属于噪音（记为 N）的概率进行更新的过程可以用如下公式表示：

$$P(N|x)=\frac{P(x|N)P(N)}{P(x)} \tag{6}$$

贝叶斯决策理论认为，个体会基于当前刺激属于信号或噪音的后验概率，来决定应当把当前刺激判断为信号还是噪音。如果刺激属于信号的后验概率大于其属于噪音的后验概率，则个体就会把当前刺激判断为信号，反之则判断为噪音(Burgess, 1985; Fleming & Daw, 2017; Lau, 2007)。由于刺激属于信号和噪音的后验概率之和为 1，因此个体会把 P (S | x) = 0.5 作为判断标准：如果刺激属于信号的后验概率大于 0.5 则判断当前刺激为信号，而如果小于 0.5 则判断刺激为噪音。

在公式(5)和(6)中，观测到刺激强度 x 的边缘概率 P (x)是相等的。具体而言，P (x)的计算方法是，根据刺激属于信号和噪音的先验概率——P (S)与 P (N)，对给定当前刺激是信号或噪音的条件下观测到 x 的概率（似然函数）——P (x | S)与 P (x | N)，进行加权平均：

$$P(x)=P(x|S)P(S)+P(x|N)P(N) \tag{7}$$

可以通过将公式(5)与公式(6)相除，消去 P (x)项，得出下面的公式：

$$\frac{P(S|x)}{P(N|x)}=\frac{P(x|S)}{P(x|N)}\cdot\frac{P(S)}{P(N)} \tag{8}$$

公式(8)意味着，贝叶斯决策过程可以被看作是个体基于后验概率的比值进行决策的过程：如果刺激属于信号和噪音的后验概率比 P (S | x) / P (N | x)大于 1，则被试会把当前刺激判断为信号，反之则判断为噪音。根据贝叶斯定理，后验概率比值是由刺激强度 x 在信号和噪音成立的条件下的似然函数比值 P (x | S) / P (x | N)，以及信号和噪音的先验概率比值 P (S) / P (N)共同决定的。

由于在信号或噪音成立的条件下，刺激强度 x 分别满足一个概率分布（通常假设为正态分布），因此 x 在信号或噪音成立条件下的似然函数反映了当刺激是信号或噪音时观测到的

刺激强度为 x 的相对概率，其等价于 x 在信号或噪音分布中出现的概率密度（即正态分布中 x 所在位置对应的纵坐标的值）。当刺激强度 x 在信号分布下出现的概率密度 $P(x \mid S)$越大，或在噪音分布下出现的概率密度 $P(x \mid N)$越小时，个体越倾向于将当前刺激判断为信号，反之则更倾向于判断为噪音。需要注意的是，个体在完成贝叶斯推理的过程中，可能并不知道信号和噪音分布的真实形态。比如，个体可能并不知道真实的信号与噪音分布的均值和标准差分别等于多少(Lau, 2007)。为了完成贝叶斯推理过程，个体必须主观推测信号和噪音分布的形态，并据此来估计当前刺激的强度 x 在信号和噪音分布中出现的似然函数(Lau, 2007)。

在一些情况下，个体对刺激属于信号还是噪音可能不存在先验的偏好。比如，被试在参加一个信号检测实验任务之前，可能对电脑屏幕上出现的刺激是信号还是噪音并没有先验的偏向，而是简单地认为二者是等可能的。此时可以简单假设信号和噪音成立的先验概率相等，即 $P(S) = P(N) = 0.5$。这种情况下的后验概率比值就等于似然函数的比值，因此贝叶斯决策的结果完全由似然函数的值来决定。然而，先验概率并不一定始终等于 0.5 (Wickens, 2001)。比如，一些个体对信号和噪音出现的先验概率可能具有一定的主观偏好，即可能偏向于认为信号或噪音其中之一更可能出现。此时，基于后验概率比值的贝叶斯决策过程不仅受到似然函数比值的影响，也受到先验概率比值的影响；当个体认为信号成立的先验概率越大，或噪音成立的先验概率越小时，就越倾向于把当前刺激判断为信号，反之则越倾向于判断为噪音。

根据贝叶斯决策理论的观点，个体在贝叶斯推理过程中使用的似然函数值和先验概率值，均来自于个体自身的主观估计。那么，当个体主观估计的似然函数和先验概率等于多少时，最终给出的决策结果是最理想的？当似然函数和先验概率的主观估计值发生变化时，个体的决策过程又会受到怎样的影响？为了回答上述问题，本文接下来将从信号检测任务中的“理想观察者”视角出发，在贝叶斯决策理论的框架下进一步探讨个体何时会给出最理想的决策，以及如何解释实际研究中被试的决策结果与理想观察者之间的偏离。

3 理想观察者

根据经典信号检测论的观点，个体会在刺激强度的数轴上设置一个报告标准 C，并通过比较当前刺激的强度 x 与 C 在数轴上的位置，来决定当前刺激是信号还是噪音。个体的作答结果可以被分为击中、虚报、漏报和正确拒绝四种类型，其中击中（将信号刺激判断为信号）和正确拒绝（将噪音刺激判断为噪音）视为正确作答，而虚报（将噪音刺激判断为信号）和漏报（将信号刺激判断为噪音）视为错误作答。那么，一个有趣的问题是，当报告标准 C 设置在哪里时，个体作答的正确率最大？在信号检测论框架下，通过设置最合适的报告标准使得作答正确率最大的个体，被称为理想观察者(ideal observer)(Wickens, 2001)。

为了计算出理想观察者会把报告标准 C 设置在怎样的位置，首先需要求出 C 值的变化如何影响作答正确率。这里以图 1 中呈现的信号检测论模型为例。在图 1 所示的模型中，信号分布和噪音分布具有相等的方差（二者均等于 1），该类模型又被称为等方差(equal variance)信号检测模型(Wickens, 2001)。根据图 1 可知，当实验中呈现的刺激为信号时，作答正确率（即击中率，hit rate，HR）等于信号分布下报告标准 C 右侧的面积。具体计算公式为：

$$HR = 1 - \Phi(C|\mu = d', \sigma = 1) \tag{9}$$

上式中，Φ (C | μ, σ)代表均值为 μ、标准差为 σ 的正态分布的累积分布函数，即等价于该正态分布位于 C 左侧的曲线下方面积。同理，当实验中呈现的刺激为噪音时，作答正确率（即正确拒绝率，correct rejection rate，CRR）等于噪音分布下报告标准 C 左侧的面积：

$$CRR = \Phi(C|\mu = 0, \sigma = 1) \tag{10}$$

假设在信号检测实验任务中，信号刺激出现的可能性为 P (S_t)，噪音刺激出现的可能性为 P (N_t)。字母 t 的意思是，上述概率代表了信号和噪音刺激在实验中出现的真实(true)概率，以与个体在贝叶斯决策过程中使用的主观先验概率相区分。据此可以求出总体的作答正确率

$P_{correct}$：

$$P_{correct} = P(S_t) \cdot HR + P(N_t) \cdot CRR = \\ P(S_t) \cdot [1 - \Phi(C|\mu = d', \sigma = 1)] + P(N_t) \cdot \Phi(C|\mu = 0, \sigma = 1) \tag{11}$$

为了求出 C 值等于多少时 $P_{correct}$ 最大，可以先求出 $P_{correct}$ 的一阶导数：

$$\frac{dP_{correct}}{dC} = P(N_t) \cdot \phi(C|\mu = 0, \sigma = 1) - P(S_t)\phi(C|\mu = d', \sigma = 1) \tag{12}$$

公式(12)中，$\phi(C\,|\,\mu, \sigma)$代表在均值为 μ、标准差为 σ 的正态分布中，刺激强度等于 C 时的概率密度。理想观察者的报告标准 C_{ideal} 应当设置在令 $P_{correct}$ 的一阶导数为 0 的位置，即满足如下的公式(13)：

$$P(N_t) \cdot \phi(C_{ideal}|\mu = 0, \sigma = 1) - P(S_t)\phi(C_{ideal}|\mu = d', \sigma = 1) = 0 \tag{13}$$

当个体把强度高于 C_{ideal} 的刺激全部判断为信号，而把强度低于 C_{ideal} 的刺激全部判断为噪音时，个体的作答正确率达到最大(Wickens, 2001)。根据正态分布的概率密度函数，可以进一步求出 C_{ideal} 在刺激强度的数轴上的位置：

$$\phi(C_{ideal}|\mu = d', \sigma = 1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(C_{ideal} - d')^2}{2}} \tag{14}$$

$$\phi(C_{ideal}|\mu = 0, \sigma = 1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{C_{ideal}^2}{2}} \tag{15}$$

$$C_{ideal} = \frac{d'}{2} + \frac{1}{d'} \ln \left[\frac{P(N_t)}{P(S_t)}\right] \tag{16}$$

在等方差信号检测模型中，当信号和噪音在实验任务中出现的概率相同，即 $P(S_t) = P(N_t) = 0.5$ 时，C 的位置恰好等于 d'/2，即理想观察者会把报告标准设置在信号分布均值和噪音分布均值的正中间位置。换言之，理想观察者设置的报告标准正好处在信号分布和噪音分布交叉的位置(Wickens, 2001)。

接下来，本文将介绍如何从贝叶斯决策理论的角度来理解理想观察者的决策规则。公式(13)又可以被写成如下的形式：

$$1 = \frac{\phi(C_{ideal}|\mu = d', \sigma = 1)P(S_t)}{\phi(C_{ideal}|\mu = 0, \sigma = 1)P(N_t)} \tag{17}$$

不难看出，公式(17)与反映贝叶斯决策过程的公式(8)非常相似。在公式(17)中，信号和

噪音在实验任务中出现的真实概率的比值 P (S_t) / P (N_t)，类似于个体在贝叶斯决策过程中使用的先验概率比值；而在公式(17)中，当刺激强度等于 C_{ideal} 时信号和噪音分布中的概率密度比值 ϕ (C_{ideal} | μ = d', σ = 1) / ϕ (C_{ideal} | μ = 0, σ = 1)，也类似于贝叶斯决策过程中的似然函数比值。事实上，贝叶斯决策理论指出，当个体完全清楚信号和噪音出现的客观先验概率以及信号和噪音分布的真实形态（即真实的似然函数值）时，采用贝叶斯决策机制可以保证个体成为理想观察者，即能够使得个体的作答正确率最大(Burgess, 1985)。具体而言，结合公式(8)与公式(17)可知，当刺激强度恰好等于 C_{ideal} 时，后验概率的比值（即信号和噪音条件下似然函数与先验概率乘积的比值）等于 1。而当刺激强度大于 C_{ideal} 时，后验概率的比值大于 1，此时个体会把当前刺激判断为信号；当刺激强度小于 C_{ideal} 时，后验概率的比值小于 1，此时个体会把当前刺激判断为噪音。上述基于贝叶斯推理过程的决策规则，与经典信号检测论框架下理想观察者的决策规则完全相同。

然而，针对实际研究的数据分析结果表明，虽然真实被试设置的报告标准与理想观察者近似(Knill, 1998; Legge et al., 2002; Stretch & Wixted, 1998)，但被试通常不会把报告标准恰好设置在公式(16)给出的 C_{ideal} 的位置，这意味着被试的决策规则与理想观察者仍然存在一定的偏离(Lau, 2007; Wickens, 2001)。根据贝叶斯决策理论，真实被试与理想观察者的偏离可

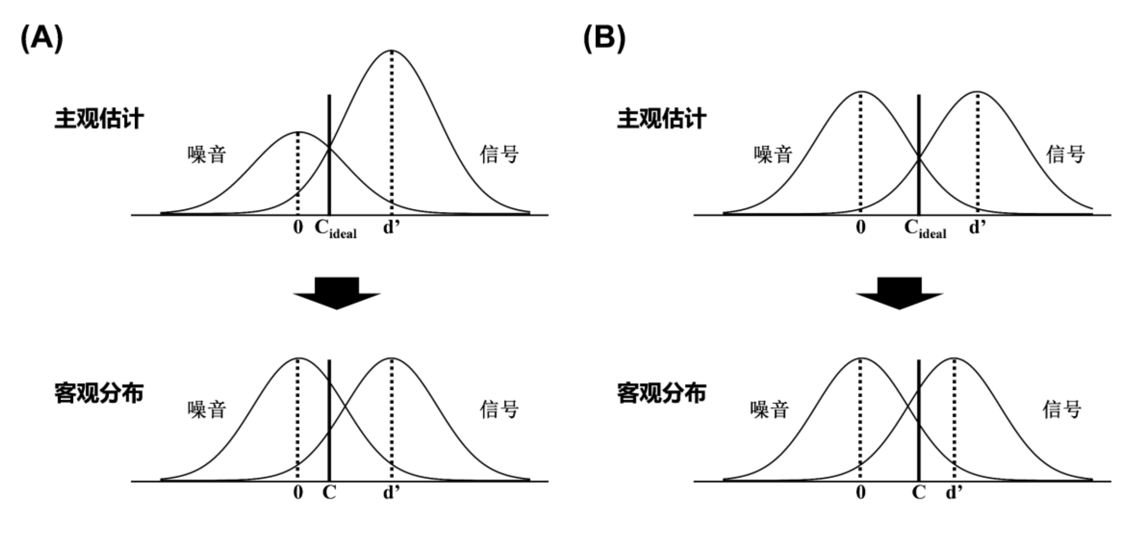


图 2 个体对先验概率和似然函数的主观估计与客观真实值之间的偏差示例

能由两种不同的原因造成。第一种原因是被试对信号和噪音出现的先验概率的估计与二者出现的实际概率不相同。图 2A 给出了一个示例：在实际的实验任务中，信号和噪音出现的客观概率相等，但被试主观认为信号出现的先验概率大于噪音的先验概率。基于贝叶斯决策理论的观点，被试此时会根据自身对先验概率的主观估计，把报告标准设置在信号分布和噪音分布均值的中间靠左的位置，以保证作答正确率最大。但由于被试主观估计的先验概率与信号和噪音出现的客观概率不同，因此被试设置的报告标准在实际的信号和噪音分布上就会与理想观察者存在偏离。第二种原因是被试对信号或噪音分布形态的主观估计与客观分布的形态不同，进而造成对似然函数的主观估计出现偏差。在图 2B 的示例中，被试对噪音分布均值的主观估计与其真实值相等，但是对信号分布均值的主观估计要大于信号分布的实际均值。在这种情况下，被试会基于贝叶斯推理的结果，把报告标准设置在信号和噪音分布均值的主观估计值的中点。然而，被试对信号分布均值的主观估计偏差会导致设置的报告标准偏离实际分布上的理想标准(Lau, 2007; Wickens, 2001)。因此，和经典信号检测论相比，贝叶斯决策理论进一步地解释了为何个体在信号检测任务中会把报告标准设置在特定的位置，并将其与个体对信号和噪音的出现概率及分布形态的主观估计紧密联系了起来。

以上结论都是基于图 1 所示的等方差信号检测模型推导得出的。在等方差信号检测模型中，虽然和经典信号检测论相比，贝叶斯决策理论可以对个体的决策过程给出更加深入的理论解释，但是二者在数学上是等价的。具体而言，虽然根据贝叶斯决策理论，个体对先验概率和似然函数的主观估计可能与其客观真实值存在差异，但只要个体主观认为信号和噪音分布的方差相等，基于经典信号检测论和贝叶斯决策理论得出的决策过程就会在数学上具有相等的形式：个体会在刺激强度 x 高于某一报告标准 C 时将刺激判断为信号，而在 x 低于 C 时将刺激判断为噪音(Glanzer et al., 2009, 2019; Wickens, 2001)。然而，在实际的信号检测任务中，信号和噪音分布的方差并不会始终相等；该类模型被称为不等方差(unequal variance)

信号检测模型(Mickes et al., 2007)。在这种情况下，经典信号检测论与贝叶斯决策理论之间的关系会变得更加复杂。接下来，本文将介绍经典信号检测论与贝叶斯决策理论在不等方差信号检测模型中存在怎样的差异。

4　不等方差信号检测模型

在信号检测论模型中，为了简化模型假设，研究者经常假设信号和噪音分布具有相等的方差（标准差）。然而，在实际研究中，这一假设并非始终成立(Green & Swets, 1966; Wixted, 2020)。比如，在前文介绍过的再认记忆任务中，研究者们发现，和“新词”代表的噪音分布相比，“旧词”代表的信号分布具有显著更大的标准差，且信号和噪音分布的标准差比值通常在 1.25:1 左右(Mickes et al., 2007; Rotello, 2017)。Wixted (2007)提出了一种假设来解释再认记忆任务中信号分布方差更大的原因：与新词相比，旧词之所以在再认记忆任务中具有更高的刺激强度，是由于个体对这些旧词在学习阶段进行了一定程度的学习与记忆；而只有当个体对所有旧词的学习程度均相等时，信号分布的方差才能与噪音分布保持一致。但实际上，个体对每个旧词的学习程度必然不是完全相等的，不同旧词的学习程度之间的变异会导致信号分布的方差显著大于噪音分布。因此，等方差信号检测模型虽然具有简单的理论假设，但可能对信号检测任务给出了过于简化的描述；而不等方差信号检测模型也许能够更好地反映个体在实际任务中完成的信号检测过程。

值得注意的是，并非所有的信号检测任务都必须使用不等方差信号检测模型来描述。比如，在二选一迫选任务(2 alternative forced choice, 2AFC)范式中，屏幕上每次会同时呈现一个信号刺激（如旧词）和一个噪音刺激（如新词），被试需要从两个刺激中准确地选出信号刺激。即使信号和噪音分布的方差不相等，2AFC 任务的信号检测过程也能够使用等方差信号检测模型来定量刻画(Wickens, 2001)。由于与本文主题不直接相关，因此本文不对 2AFC

任务下的信号检测论模型进行进一步介绍。感兴趣的读者可以阅读已有文献中对 2AFC 任务的论述(Macmillan & Creelman, 2004; Wickens, 2001)。

下面，本文将介绍在不等方差信号检测模型的框架下，个体如何根据贝叶斯决策理论来完成决策过程。图 3 给出了不等方差信号检测模型的示例。与图 1 相似，图 3 中噪音分布的均值为 0，信号分布的均值为 d'。然而，与图 1 不同的是，图 3 中噪音分布的标准差 σ_N 等于 1，而信号分布的标准差 σ_S 与 1 并不相等（图 3 中假设 σ_S 大于 1，即与再认记忆任务中的情况相同）。为了简化贝叶斯推理过程，这里假设信号和噪音出现的真实概率均为 0.5，且个体完全清楚信号和噪音的先验概率及似然函数。由于信号和噪音出现的先验概率相等，因此个体的决策结果完全依赖于当前刺激的强度 *x* 在信号和噪音分布中的似然函数的比值：根据公式(8)，当似然函数的比值大于 1 时，个体会将当前刺激判断为信号，反之则判断为噪音。然而，根据图 3 可知，在不等方差信号检测模型中，信号分布和噪音分布存在两个不同位置的交叉点（C_1 和 C_2）。当 *x* 小于 C_1 或大于 C_2 时，*x* 在信号分布中的概率密度高于噪音

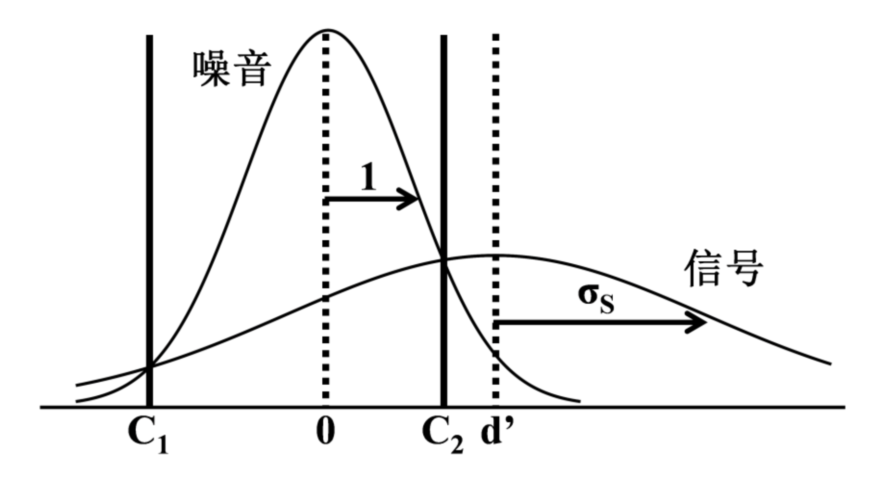


图 3 不等方差信号检测模型示例

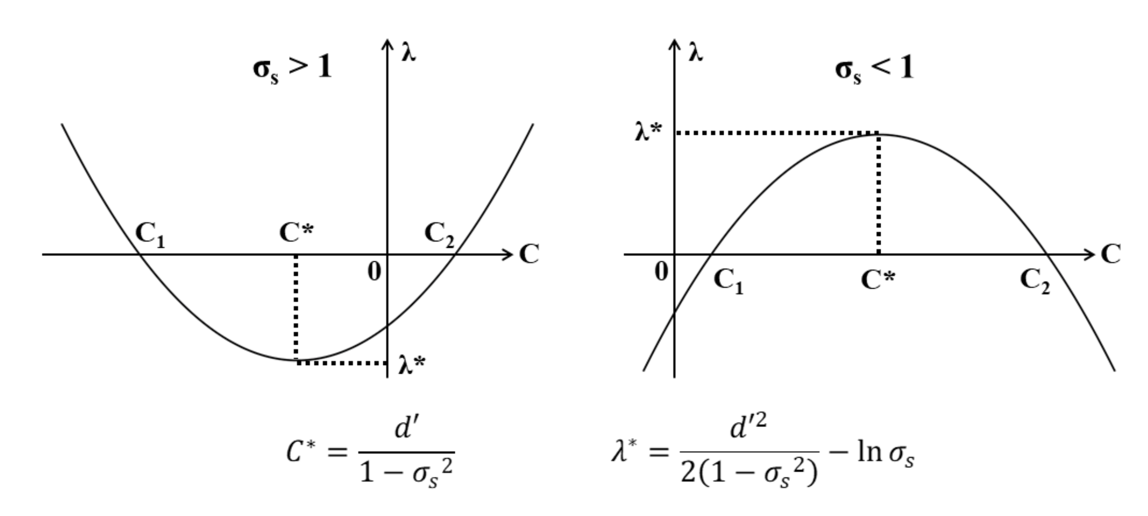
分布，因此被试会把当前刺激判断为信号；而当 *x* 处在 C_1 和 C_2 之间时，*x* 在噪音分布中的概率密度更高，被试会把当前刺激判断为噪音。换言之，信号和噪音分布的似然函数的一个特定比值，在刺激强度的数轴上对应两个不同的报告标准。

Glanzer 等人(2009)通过严密的数学推导，给出了在不等方差信号检测模型中，似然函数

比值与刺激强度数轴上的报告标准 C 之间的对应关系。Glanzer 等人首先计算了当刺激强度等于 C 时，信号和噪音分布中似然函数的比值，随后对似然函数比值取自然对数（记为 λ）：

$$\lambda = \ln\left[\frac{\phi(C|\mu = d', \sigma = \sigma_S)}{\phi(C|\mu = 0, \sigma = 1)}\right] \tag{18}$$

由于对数函数是单调递增的，因此当 λ 取到其最大值（或最小值）时，似然函数的比值也应当能够取到最大值（或最小值）。进一步的计算结果表明(Glanzer et al., 2009)，λ 与刺激强度数轴上的报告标准 C 之间满足二次函数关系（见图 4）。当信号分布的标准差 σ_S 大于 1 时，λ 有最小值（记为 λ*）；而当 σ_S 小于 1 时，λ 有最大值（记为 λ*）。当且仅当 λ 等于 λ* 时，刺激强度数轴上存在唯一的报告标准与其对应（记为 C*）；而当 λ 不等于 λ*时，每一个 λ 值都对应两个不同的报告标准。当 λ 等于 0（即似然函数比值为 1）时，对应的两个报告标准 C_1 和 C_2 将整个刺激强度数轴分为三个部分。如果 σ_S 大于 1，则个体会将 C_1 和 C_2 中间的部分判断为噪音，而将两侧的部分判断为信号；如果 σ_S 小于 1 则相反。



$$C^* = \frac{d'}{1 - {\sigma_S}^2} \qquad \lambda^* = \frac{d'^2}{2(1 - {\sigma_S}^2)} - \ln \sigma_S$$

图 4 不等方差信号检测模型中 λ 与 C 的关系

另一方面，经典信号检测论认为，即使在不等方差信号检测模型中，个体仍然只会简单地设置一个报告标准 C，并当刺激强度 x 高于 C 时将刺激判断为信号，反之则判断为噪音(Mickes et al., 2007; Wickens, 2001)。因此，经典信号检测论和贝叶斯决策理论为不等方差信号检测模型中的决策机制给出了不同的预测，个体根据两种理论给出的决策结果也不再完全

等价（这一点与等方差信号检测模型有别）。

经典信号检测论和贝叶斯决策理论在不等方差信号检测模型中存在的差异可以从操作者特性曲线(receiver operating characteristic curve, ROC curve)中体现出来。ROC 曲线指的是在保持信号和噪音的客观分布形态不变的基础上，反映击中率(HR)和虚报率(false alarm rate, FAR)之间关系的曲线(Wickens, 2001)。在经典信号检测论框架下，通过改变报告标准 C 的位置（如设置宽松或严格的报告标准），可以同时改变击中率和虚报率的值。而在贝叶斯决策理论框架下（假设个体完全清楚先验概率和似然函数的真实值），可以通过改变信号和噪音刺激出现概率的方式来影响先验概率，进而影响后验概率比值等于 1 时的似然函数比值（即影响 λ 的临界值）；λ 的临界值对应的报告标准的位置也会随之发生变化，进而改变击中率和虚报率。将所有可能的击中率和虚报率配对绘制在图上，以虚报率为横坐标，击中率为纵坐标，即可得到 ROC 曲线。

在等方差信号检测模型中，无论个体基于经典信号检测论还是贝叶斯决策理论来完成决策过程（二者在数学上是等价的），只要辨别力指数 d’的值大于 0，ROC 曲线就会始终处在对角线上方，即意味着击中率始终高于虚报率；当 d’的值越高时，ROC 曲线下的面积越大（见图 5A 的示例）。但是，在不等方差信号检测模型中，如果个体基于经典信号检测论给出的决策机制来完成决策过程，则就算 d’的值大于 0，ROC 曲线也会有一部分处在对角线下方。比如，当信号分布标准差 σ_S 大于 1 时，在击中率和虚报率都非常接近于 1 的情况下，击中率会小于虚报率（见图 5B）；而当 σ_S 小于 1 时，则在击中率和虚报率都接近于 0 的情况下，击中率小于虚报率（见图 5C）。与之相对，如果个体基于贝叶斯决策理论来完成决策过程，则即使在不等方差信号检测模型中，ROC 曲线也会始终处在对角线上方（见图 5D）；这一结论可以在数学上得到严格的证明(Wickens, 2001)。

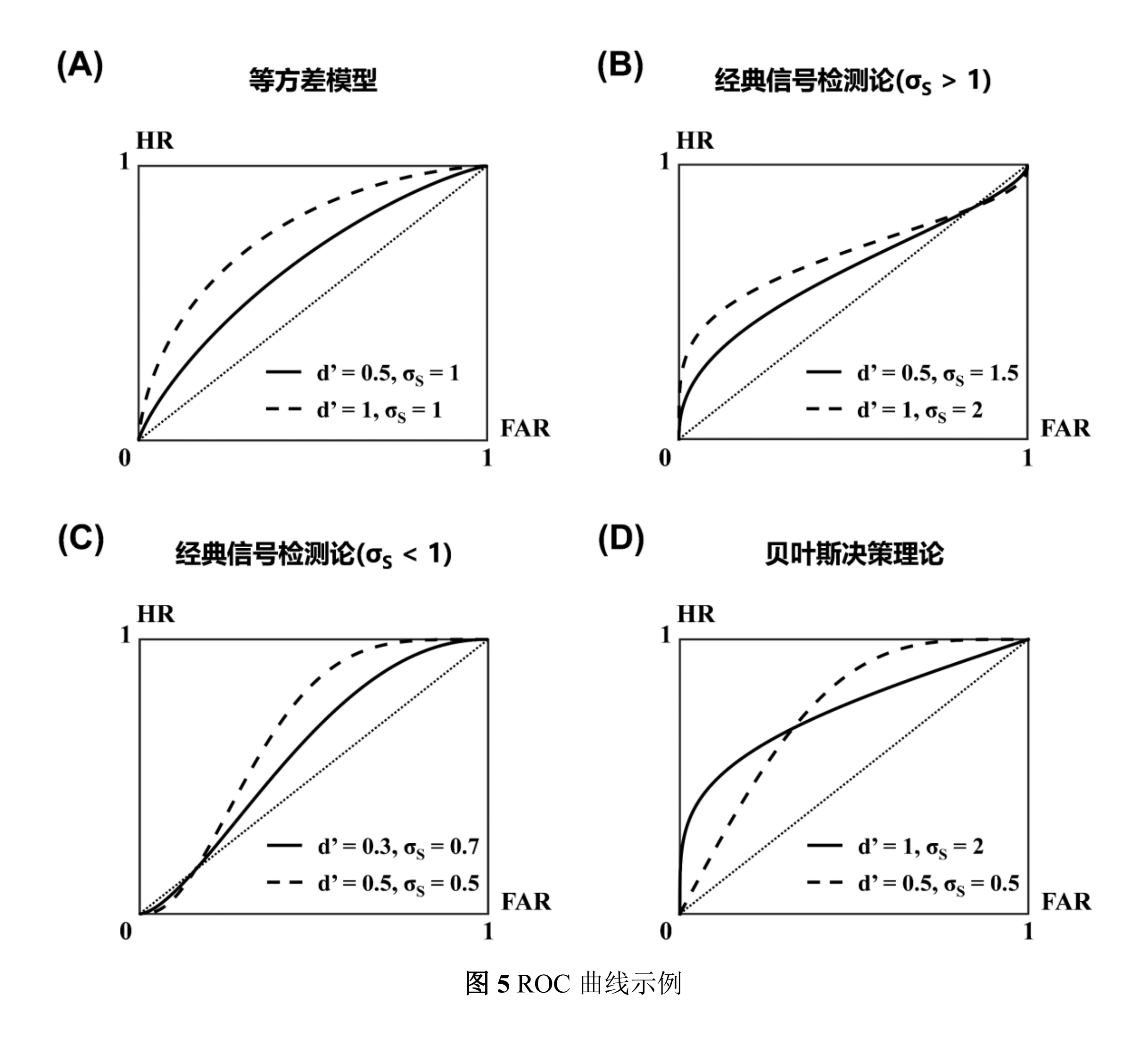


图 5 ROC 曲线示例

经典信号检测论和贝叶斯决策理论之间的差异可以在 zROC 曲线上得到更加直观的体现(Macmillan & Creelman, 2004)。所谓 zROC 曲线，指的是首先利用下面的公式将击中率 HR 和虚报率 FAR 转化为 z 分数：

$$zHR = \Phi^{-1}(HR|\mu = 0, \sigma = 1) \quad (19)$$

$$zFAR = \Phi^{-1}(FAR|\mu = 0, \sigma = 1) \quad (20)$$

公式(19)和(20)中，$\Phi^{-1}(p\,|\,\mu = 0, \sigma = 1)$代表均值为 0、标准差为 1 的标准正态分布累积分布函数的逆函数，即将取值范围在 0 到 1 之间的概率值 p 转化为取值范围在负无穷到正无穷之间的 z 分数。随后，以 zFAR 为横坐标，zHR 为纵坐标，绘制二者之间的关系图，即为 zROC 曲线。图 6 给出了 zROC 曲线的示例。当 zROC 曲线位于直线 $y = x$ 的上方时，就意味着击中率大于虚报率。从数学上可以证明，在经典信号检测论框架下，zHR 和 zFAR 之间

的关系为一条直线，其斜率等于噪音分布标准差和信号分布标准差的比值 σ_N/σ_S (Wickens, 2001)。在信号检测论中，噪音分布标准差 σ_N 通常设为 1。当信号分布标准差 σ_S 大于 1 时，zROC 直线的斜率小于 1，此时 zROC 直线必然会在右侧与直线 $y=x$ 相交，并导致当击中率和虚报率的值均很大时，击中率小于虚报率。同理，当 σ_S 小于 1 时，zROC 直线的斜率大于 1，会在左侧与 $y=x$ 相交，并导致当击中率和虚报率的值均很小时，击中率小于虚报率（见图 6A 的示例）。另一方面，在贝叶斯决策理论的框架下，zHR 和 zFAR 之间的关系为一条曲线；无论 σ_S 的值等于多少，zROC 曲线始终位于 $y=x$ 的上方（见图 6B 的示例）(Macmillan & Creelman, 2004)。

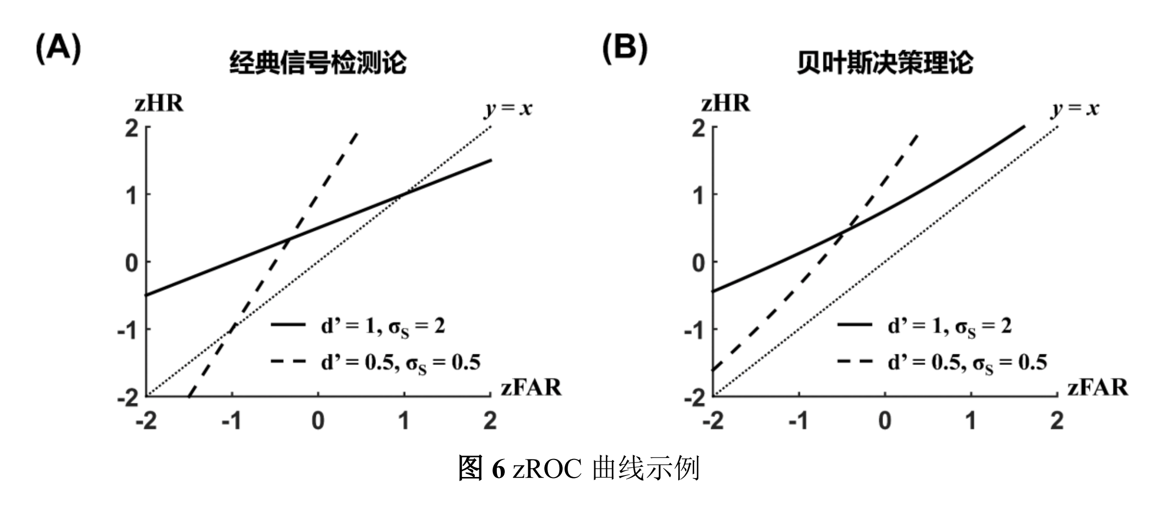


图 6 zROC 曲线示例

由于在不等方差信号检测模型中，经典信号检测论和贝叶斯决策理论预测了不同的 ROC 曲线和 zROC 曲线形态，因此从理论上讲，可以通过考察实际研究当中获得的 ROC 以及 zROC 曲线，以及在极端条件下获得的击中率和虚报率数据，来确认哪一种理论能够更好地解释真实被试在不等方差信号检测任务当中的决策过程。比如，在再认记忆任务中，可以考虑通过尽可能地增加旧词的出现频率，降低新词的出现频率，使得被试设置尽可能宽松的报告标准，即将绝大多数刺激都判断为旧词（此时击中率和虚报率都接近于 1）。在该实验条件下，如果发现被试的击中率低于虚报率，则支持经典信号检测论；如果击中率始终高于虚报率，则支持贝叶斯决策理论。

然而，上述的实验设计很难在实际研究中得到应用。这是因为在经典信号检测论框架下，当不等方差信号检测模型中的击中率低于虚报率时，击中率与虚报率的真值都非常接近于极端值 0 或 1。同时，即使存在击中率低于虚报率的情况，二者之间的差异也是很小的（见图 5B 和图 5C）。在此情况下，样本数据中计算得到的击中率与虚报率差值会受到抽样误差的严重影响，一点细微的抽样误差的存在都会导致研究者难以根据实验结果准确估计击中率与虚报率的真实差异(Glanzer et al., 2019; Macmillan & Creelman, 2004)。那么，是否存在某种实验设计，使得研究者可以比较经典信号检测论和贝叶斯决策理论在解释真实个体的决策过程时的优劣呢？

5　支持贝叶斯决策理论的研究证据

有研究者指出，两条件(two-condition)实验可以被用于比较经典信号检测论和贝叶斯决策理论在真实决策过程中的合理性(Glanzer et al., 2009, 2019; Semmler et al., 2018; Stretch & Wixted, 1998)。所谓的两条件实验，指的是在一个被试内实验设计中，要求同一批被试完成两个不同的信号检测任务；这两个任务除了存在难度上的区别外，在任务要求上没有任何差异。比如，被试可以完成两个不同的再认记忆任务，这两个任务的唯一区别在于学习阶段被试学习每个旧词的时间长度存在差异，学习时间越长则任务难度越低(Glanzer et al., 2009)。同时，在两个信号检测任务中，被试不仅需要判断每个刺激是信号还是噪音，而且需要给出自己对作答结果的信心判断(confidence rating)，即被试需要在一个李克特量表上评价自己对作答结果的信心程度如何。此外，Glanzer 等人(2019)指出，在两条件实验设计中，被试必须清楚意识到每个 trial 来自哪个任务难度条件。因此，两个不同难度的信号检测任务最好在不同的组块(block)中进行；如果一定要使用混合列表设计(mixed-list design)，即把两个难度的信号检测任务的 trial 混合在一起，则必须通过特定的提示告知被试当前的 trial 来自哪个

实验条件，比如为屏幕上呈现的词语标注不同的颜色等。

研究者们提出了似然比理论(likelihood ratio theory)来解释个体在两条件实验中的决策过程(Glanzer et al., 2009, 2019; Semmler et al., 2018)。似然比理论以贝叶斯决策理论为基础，包含两个具体假设：（1）似然比不变假设。该假设认为，在两种不同难度的实验条件下，个体用于贝叶斯决策过程的似然函数比值的临界值保持不变（这一假设会在下面进行更详细的介绍）。（2）真实似然比假设。该假设认为，个体知道信号和噪音分布的真实分布形态，并会依据刺激强度在信号和噪音分布中的真实似然函数比值来完成决策过程。需要注意的是，真实似然比假设是比一般的贝叶斯决策理论更强的理论假设：贝叶斯决策理论通常允许个体主观估计的似然函数比值与真实的似然函数比值存在差异(Lau, 2007; Wickens, 2001)，但真实似然比假设则认为个体知道（或至少近似知道）信号与噪音分布中的真实似然函数(Glanzer et al., 2019; Semmler et al., 2018)。根据似然比理论，个体在两条件实验中的决策过程应同时满足上述两个假设(Glanzer et al., 2009, 2019)。

在似然比理论的基础上，Glanzer 等人(2009, 2019)预测了两条件实验中可能出现的三种实验现象，包括镜像效应(mirror effect)、方差效应(variance effect)和 zROC 长度效应(zROC length effect)。Glanzer 等人指出，似然比理论可以很容易地解释上述三种现象的产生机制，但经典信号检测理论难以同时解释这三种现象的发生。实际研究结果表明，在不同认知领域（如知觉、记忆、推理、心理旋转等）的信号检测任务中，上述三种现象均稳定存在(Glanzer et al., 2009, 2019; Hilford et al., 2015, 2019; Semmler et al., 2018)，即在一定程度上支持了贝叶斯决策理论的合理性。由于篇幅所限，本文只着重介绍其中的一种现象——zROC 长度效应。

zROC 长度效应最早由 Stretch 和 Wixted (1998)发现。Stretch 和 Wixted 对已有的一项研究中的数据进行了再次分析(Ratcliff et al., 1994)；在该研究中，被试需要完成两种不同难度的再认记忆任务（容易 vs.困难）。比如，在容易条件下被试学习每个旧词的时间为 3 s，而

在困难条件下每个旧词的学习时间只有 1 s。在再认测验中，被试需要在一个 6 点量表上针对每个词语完成判断任务，其中 1 到 6 分别代表“确定是新词”、“很可能是新词”、“可能是新词”、“可能是旧词”、“很可能是旧词”、“确定是旧词”。因此，被试不仅需要针对每个词语进行新旧判断，还需要报告自己对作答结果的信心程度。在经典信号检测论框架下，信心判断任务是二分判断任务（信号 vs.噪音）的推广形式，即在 *n* 点量表上进行的信心判断任务等价于在刺激强度的数轴上设置(*n*-1)个报告标准(Wickens, 2001)。如图 7A 所示，当刺激强度低于第一个报告标准 C_1 时，被试会在信心判断量表上报告为 1；当刺激强度在 C_1 和 C_2 之间时，被试会在信心判断量表上报告为 2；以此类推。Stretch 和 Wixted 发现，与困难条件（辨别力指数 d’更低的条件）相比，在容易条件（d’更高的条件）下，所有的报告标准的位置均会向中间收缩，即不同报告标准之间的距离会缩短（见图 7A）。Stretch 和 Wixted 随后在一个新的实验研究中验证了上述结论。

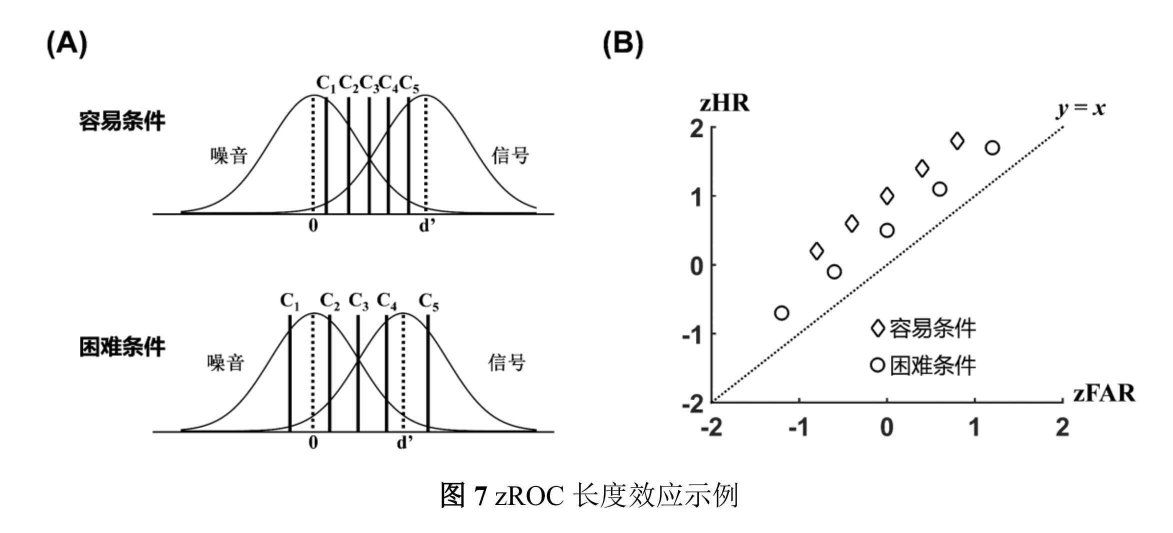


图 7 zROC 长度效应示例

Glanzer 等人(2009, 2019)使用 zROC 曲线对 Stretch 和 Wixted (1998)报告的现象进行了重新表述。具体而言，Glanzer 等人首先根据被试在 *n* 点信心判断量表上给出的判断结果，通过改变报告标准的方式计算出(*n*-1)对击中率和虚报率。比如，在一个 6 点信心判断量表上，假设将信心判断量表点数 1 视为被试把当前刺激判断为噪音，而将量表点数 2-6 视为被试把刺激判断为信号，即可计算出一组击中率和虚报率；而如果将量表点数 1-2 视为被试把

刺激判断为噪音，而将量表点数 3-6 视为被试把刺激判断为信号，则可计算出另一组击中率和虚报率；以此类推。随后，Glanzer 等人基于每种实验条件下获得的(n-1)对击中率和虚报率绘制出 zROC 曲线。结果表明，与困难条件相比；容易条件下绘制的 zROC 曲线两端之间的距离显著更短（如图 7B 所示）；这意味着当改变报告标准时，容易条件下的击中率和虚报率值变化幅度更小，即说明容易条件下不同的报告标准之间更为接近。Glanzer 等人将这一现象命名为 zROC 长度效应。

似然比理论指出，基于似然比不变假设可以很自然地推导出 zROC 长度效应。根据贝叶斯决策理论（参考公式(8)），个体在 n 点信心判断量表上给出判断的过程，可以被看作是在后验概率比值上设置(n-1)个阈值（记为 β_1、β_2……β_{n-1}），并将当前刺激对应的后验概率比值与每个 β 值相比较的过程。假如当前刺激对应的后验概率比值小于 β_1，则个体在信心判断量表上报告的结果为 1；假如刺激对应的后验概率比值在 β_1 和 β_2 之间，则信心判断的结果为 2；以此类推(Green & Swets, 1966)。Glanzer 等人(2009, 2019)假设，当两个实验条件之间的唯一差异为任务难度时，个体设置的阈值 β 应当不会在实验条件之间发生改变。由于后验概率比值等于先验概率比值和似然函数比值的乘积，且个体在贝叶斯推理过程中对于信号和噪音出现的先验概率的主观估计通常可以近似视为恒定(Fleming & Daw, 2017; Wickens, 2001)，因此可以在数学上等价地认为，个体在信号和噪音分布的似然函数比值上设置了一组阈值，并通过每个刺激对应的似然函数比值与上述阈值之间的比较来给出信心判断的结果；同时，个体在似然函数比值上设置的阈值应当不会在两个实验条件之间发生变化(Glanzer et al., 2019)。上述假设即为似然比不变假设。

接下来，本文将以等方差信号检测模型为例，根据似然比不变假设来推导 zROC 长度效应。由于个体在似然函数比值上设置的阈值是恒定不变的，因此在为似然函数比值取对数之后，对数似然比（λ）上设置的阈值也不会在实验条件之间发生改变。基于正态分布的概率

密度函数（参考公式(14)和(15)），可以写出 λ 与报告标准 C 之间的对应关系(Glanzer et al., 2009)：

$$\lambda = \ln\left[\frac{\phi(C|\mu = d', \sigma = 1)}{\phi(C|\mu = 0, \sigma = 1)}\right] = d'C - \frac{d'^2}{2} \tag{21}$$

$$C = \frac{\lambda}{d'} + \frac{d'}{2} \tag{22}$$

当 λ 在不同实验条件之间维持恒定时，随着任务难度的下降（即 d'的上升），报告标准 C 会逐渐向 d'/2 的周围收缩，因此一组报告标准之间的距离会缩短，即得出了 zROC 长度效应。此外，Glanzer 等人(2009)的数学推导结果表明，在不等方差信号检测模型中，虽然个体的贝叶斯决策机制更为复杂，但只要个体保持 λ 在不同实验条件之间的恒定， zROC 长度效应就会成立。Glanzer 等人(2009, 2019)指出，虽然经典信号检测论也可以对 zROC 长度效应进行强行解释，即认为个体在刺激强度的数轴上直接设置的报告标准恰好随着任务难度的降低而收缩，但它无法解释报告标准的收缩对应怎样的内在心理机制。因此，zROC 长度效应可以被视为是支持贝叶斯决策理论的证据。基于相似的方法，Glanzer 等人(2009)从似然比理论中推导出了另外两种现象——镜像效应和方差效应，并在实际研究结果中验证了这两种效应的存在。感兴趣的读者可以阅读 Glanzer 等人(2009)的数学推导过程。

根据似然比理论得出的三种实验现象在一定程度上为贝叶斯决策理论提供了支持。然而，似然比理论一直受人质疑的地方在于其包含的真实似然比假设，即认为个体在信号检测任务中能够知道信号和噪音分布的真实形态。一些研究者质疑真实被试是否有能力掌握如此多的信息(Balakrishnan & Ratcliff, 1996; Criss & McClelland, 2006)。为了回应这一质疑，Semmler 等人(2018)提出，个体对信号和噪音分布形态的认识来自于日常生活中经年累月的学习；即使个体不能通过学习来完全精确地掌握信号与噪音分布的均值与标准差，也至少可以清楚地意识到信号和噪音分布之间的距离如何随着任务难度的变化而发生变化(Turner et al., 2011; Wixted & Gaitan, 2002)。然而，上述观点目前仍主要停留在理论层面。未来需要进一步研究

对似然比理论的假设合理性进行考察。

6 结语

本文在贝叶斯决策理论的框架下，探讨了个体如何通过贝叶斯推理过程完成信号检测任务中的决策判断。与经典信号检测论相比，贝叶斯决策理论可以更加深入地解释个体辨别信号和噪音的决策过程对应怎样的内在心理机制。但是，贝叶斯决策理论中涉及的参数较多（包括个体对先验概率和似然函数的主观估计），而在信号检测实验任务中，根据击中率和虚报率的实验数据难以对上述的所有参数进行求解。比如，个体在信号检测任务中的报告标准发生变化，既有可能是源于个体改变了对先验概率的主观估计，也可能是由于对似然函数的主观估计发生改变；而在实际研究中很难对上述两种情况进行区分(Fleming & Daw, 2017; Lau, 2007)。虽然似然比理论认为个体对似然函数的主观估计和客观真实值基本相同(Glanzer et al., 2009, 2019; Semmler et al., 2018)，即对贝叶斯决策理论进行了简化，但这一假设受到了其他研究者的质疑。因此，本文认为，在实际的信号检测实验任务中，经典信号检测论（以及该理论提出的报告标准 C）依然是可用于数据分析的有效工具。

本文讨论的贝叶斯决策理论并不能完整反映个体在信号检测任务中的决策机制。比如，贝叶斯决策理论认为，个体的决策结果完全基于信号和噪音出现的后验概率的比值，当信号的后验概率大于噪音时，被试会将当前刺激判断为信号，反之则判断为噪音(Burgess, 1985)。然而，在实际研究中，被试的决策结果还会受到与任务相关的奖励及惩罚程度的影响。比如，一名司机在驾驶汽车即将到达十字路口时，需要快速判断前方的红绿灯上目前显示的是哪种颜色（这里定义绿灯为信号，红灯为噪音）。在上述的知觉判断任务中，漏报（即将绿灯判断为红灯）并不会产生非常严重的后果，只会导致司机在十字路口多等待一段时间。然而，虚报（将红灯判断为绿灯）则会造成严重危害，比如司机闯红灯撞到绿灯方向通行的汽车或

行人。因此，司机可能会采取非常严格的报告标准，即倾向于将红绿灯上显示的颜色判断为红色（噪音），而这一判别标准与司机实际接收的知觉信息（刺激强度）是相互独立的。Green 和 Swets (1966)指出，当不同作答结果的奖励或惩罚程度不同时，个体的作答目标可能会从最大化作答正确率（即理想观察者），转变为最大化自身获得的奖励或最小化自身所受的惩罚。此时，个体的最终决策会同时基于信号和噪音成立的后验概率，以及不同的作答反应可能导致的后果。

前文提到，贝叶斯决策理论认为，个体在信号检测任务中给出信心判断的过程，可以被视为是在后验概率比值上设置一组不同的阈限，并将当前刺激对应的后验概率比值与每个阈限值相比较的过程(Glanzer et al., 2009, 2019; Green & Swets, 1966)。然而，近年来，研究者逐渐认识到，基于刺激强度推理得到的后验概率并不能准确地反映个体的信心判断结果。这是因为个体在决策当前刺激是信号还是噪音时依赖的刺激强度，与信心判断过程中使用的信息可能并不完全相同：个体在信心判断中可能会丢失一部分关于刺激内容的信息，也可能基于一些额外的推理过程获得更丰富的信息(Hu et al., 2021; Jang et al., 2012; Maniscalco & Lau, 2012; Shekhar & Rahnev, 2021)。为了将信心判断过程更好地纳入到贝叶斯决策理论的框架下，Fleming 和 Daw (2017) 对传统的贝叶斯决策理论进行了改进，并在改进后的理论中考虑了个体在决策过程和信心判断过程中利用信息的差异。但是，另一些研究者认为，即使考虑了个体在信心判断过程中利用的独特信息，信心判断过程自身的内在机制可能也与贝叶斯决策理论的观点并不完全相符(Adler & Ma, 2018; Li & Ma, 2020)。关于信心判断对应的内在心理过程是否为贝叶斯决策过程，目前仍然是一个存在争论的话题，需要未来研究的进一步探索。

参考文献

胡传鹏, 孔祥祯, Wagenmakers, E.-J., Ly, A., 彭凯平. (2018). 贝叶斯因子及其在JASP中的实现. *心理科学进展*, *26*(6), 951–965.

Adler, W. T., & Ma, W. J. (2018). Comparing Bayesian and non-Bayesian accounts of human confidence reports. *PLoS Computational Biology*, *14*(11), e1006572.

Balakrishnan, J. D., & Ratcliff, R. (1996). Testing models of decision making using confidence ratings in classification. *Journal of Experimental Psychology: Human Perception and Performance*, *22*(3), 615–633.

Banks, W. P. (1970). Signal detection theory and human memory. *Psychological Bulletin*, *74*(2), 81–99.

Burgess, A. (1985). Visual signal detection.III.On Bayesian use of prior knowledge and cross correlation. *Journal of the Optical Society of America A*, *2*(9), 1498–1507.

Criss, A. H., & McClelland, J. L. (2006). Differentiating the differentiation models: A comparison of the retrieving effectively from memory model (REM) and the subjective likelihood model (SLiM). *Journal of Memory and Language*, *55*(4), 447–460.

Egan, J., Schulman, A. I., & Greenberg, G. Z. (1959). Operating characteristics determined by binary decisions and by ratings. *The Journal of the Acoustical Society of America*, *31*(6), 768–773.

Fleming, S. M., & Daw, N. D. (2017). Self-evaluation of decision-making: A general bayesian framework for metacognitive computation. *Psychological Review*, *124*(1), 91–114.

Glanzer, M., Hilford, A., Kim, K., & Maloney, L. T. (2019). Generality of likelihood ratio decisions. *Cognition*, *191*, 103931.

Glanzer, M., Hilford, A., & Maloney, L. T. (2009). Likelihood ratio decisions in memory: Three

implied regularities. *Psychonomic Bulletin & Review*, *16*(3), 431–455.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. Wiley.

Hilford, A., Glanzer, M., Kim, K., & Maloney, L. T. (2019). One mirror effect: The regularities of recognition memory. *Memory & Cognition*, *47*(2), 266–278.

Hilford, A., Maloney, L. T., Glanzer, M., & Kim, K. (2015). Three regularities of recognition memory: the role of bias. *Psychonomic Bulletin & Review*, *22*(6), 1646–1664.

Hu, X., Zheng, J., Su, N., Fan, T., Yang, C., Yin, Y., Fleming, S. M., & Luo, L. (2021). A Bayesian inference model for metamemory. *Psychological Review*, *128*(5), 824–855.

Jang, Y., Wallsten, T. S., & Huber, D. E. (2012). A stochastic detection and retrieval model for the study of metacognition. *Psychological Review*, *119*(1), 186–200.

Kersten, D., Mamassian, P., & Yuille, A. (2004). Object Perception as Bayesian Inference. *Annual Review of Psychology*, *55*(1), 271–304.

Knill, D. C. (1998). Discrimination of planar surface slant from texture: human and ideal observers compared. *Vision Research*, *38*(11), 1683–1711.

Lau, H. C. (2007). A higher order Bayesian decision theory of consciousness. *Progress in Brain Research*, *168*, 35–48.

Legge, G. E., Hooven, T. A., Klitz, T. S., Stephen Mansfield, J., & Tjan, B. S. (2002). Mr. Chips 2002: new insights from an ideal-observer model of reading. *Vision Research*, *42*(18), 2219–2234.

Li, H.-H., & Ma, W. J. (2020). Confidence reports in decision-making with multiple alternatives violate the Bayesian confidence hypothesis. *Nature Communications*, *11*(1), 2004.

Macmillan, N. A., & Creelman, C. D. (2004). *Detection theory: A user's guide*. London: Psychology press.

Maloney, L. T., & Zhang, H. (2010). Decision-theoretic models of visual perception and action. *Vision Research*, *50*(23), 2362–2374.

Mamassian, P. (2016). Visual confidence. *Annual Review of Vision Science*, *2*(1), 459–481.

Maniscalco, B., & Lau, H. (2012). A signal detection theoretic approach for estimating metacognitive sensitivity from confidence ratings. *Consciousness and Cognition*, *21*(1), 422–430.

Mickes, L., Wixted, J. T., & Wais, P. E. (2007). A direct test of the unequal-variance signal detection model of recognition memory. *Psychonomic Bulletin & Review*, *14*(5), 858–865.

Pouget, A., Drugowitsch, J., & Kepecs, A. (2016). Confidence and certainty: Distinct probabilistic quantities for different goals. *Nature Neuroscience*, *19*(3), 366–374.

Ratcliff, R., McKoon, G., & Tindall, M. (1994). Empirical generality of data from recognition memory receiver-operating characteristic functions and implications for the global memory models. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(4), 763–785.

Rotello, C. M. (2017). Signal Detection Theories of Recognition Memory. In J. Byrne (Ed.), *Learning and Memory: A Comprehensive Reference (Second Edition)* (pp. 201–225). Academic Press.

Semmler, C., Dunn, J., Mickes, L., & Wixted, J. T. (2018). The role of estimator variables in eyewitness identification. *Journal of Experimental Psychology: Applied*, *24*(3), 400–415.

Shekhar, M., & Rahnev, D. (2021). The nature of metacognitive inefficiency in perceptual decision making. *Psychological Review*, *128*(1), 45–70.

Stretch, V., & Wixted, J. T. (1998). Decision rules for recognition memory confidence judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*(6), 1397–1410.

Tanner, W. P., & Swets, J. A. (1954). A decision-making theory of visual detection. *Psychological Review*, *61*(6), 401–409.

Turner, B. M., Van Zandt, T., & Brown, S. (2011). A dynamic stimulus-driven model of signal detection. *Psychological Review*, *118*(4), 583–613.

Wickens, T. D. (2001). *Elementary signal detection theory*. New York: Oxford university press.

Wixted, J. T. (2007). Dual-process theory and signal-detection theory of recognition memory. *Psychological Review*, *114*(1), 152–176.

Wixted, J. T. (2020). The forgotten history of signal detection theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *46*(2), 201–233.

Wixted, J. T., & Gaitan, S. C. (2002). Cognitive theories as reinforcement history surrogates: The case of likelihood ratio models of human recognition memory. *Animal Learning & Behavior*, *30*(4), 289–305.

The Relationship Between Signal Detection Theory and Bayesian Decision Theory

HU Xiao

(Faculty of Psychology, Beijing Normal University, Beijing 100875)

Abstract: Signal detection theory (SDT) has been widely applied to explain the decision-making process in different types of cognitive tasks. However, one important limitation of classical SDT is that it is difficult to illustrate the mental mechanisms underlying the setting of response criterion. The current article discusses the decision rule in signal detection tasks from the perspective of Bayesian decision theory (BDT). I first introduce the basic idea of BDT based on the Bayes' theorem. Next, I discuss how BDT explains the decision rule of ideal observer, and characterizes the deviation between actual participants and ideal observer in empirical signal detection tasks. I then discuss the difference between classical SDT and BDT in unequal variance signal detection model. Finally, I briefly introduce the empirical research evidence supporting BDT.

Key words: signal detection theory; Bayesian decision theory; prior probability; likelihood function; response criterion